DOCUMENT RESUME

ED 299 231 SP 030 597

AUTHOR Redinger, Linda
TITLE Evaluation: A Means of Improving Teacher Performance.
PUB DATE Apr 88
NOTE 41p.; Exit Paper, University of Indiana of South Bend.
PUB TYPE Reference Materials - Bibliographies (131) -- Dissertations/Theses - Undetermined (040) -- Reports - Research/Technical (143)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Elementary Secondary Education; *Evaluation Criteria; *Evaluation Methods; Observation; Teacher Attitudes; *Teacher Behavior; Teacher Effectiveness; *Teacher Evaluation; *Teacher Improvement

ABSTRACT

Abstracts are presented of research studies on various methods of teacher evaluation. The purpose of this study was to seek out those factors in evaluation programs that are necessary for enhancing teacher effectiveness. The research studies are organized by four categories: (1) status of teacher evaluation; (2) attributes of successful evaluation programs; (3) problems with typical evaluation programs; and (4) case studies of successful research-based programs. A summary offers an analysis of the types of evaluation programs reviewed and some recommendations are made. A list of references is included. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED299231

EVALUATION: A MEANS OF IMPROVING TEACHER PERFORMANCE

By

LINDA REDINGER
Graduate Student

Study Directed
by
Charles R. DuVall, Ph.D.

Submitted in partial fulfillment of the
requirements of the course S591---
exit requirements IUSB

April 1988

SP 030 597

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

L. Redinger

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

BEST COPY AVAILABLE

CHAPTER I

Evaluation of teacher performance is a process dreaded by many teachers and administrators and ridiculed by others. As one teacher explains:

> I could care less what the principal says. He doesn't spend enough time in my classroom to know what is going on (McCarty 352).

For the majority of teachers, the typical evaluation involves a single in-class visit once a year or less (McCarty 352). Poor assessments involve evaluators who are concerned with neat rooms or bulletin boards rather than giving reasons for teachers' ratings or suggestions for improvement (Turner 40). Teachers often feel that evaluation has a negative effect or no effect on their teaching (Turner 40).

Over the past decade, teacher evaluation has become a national issue. Evaluation systems which measure teacher effectiveness and reward teacher excellence have been recommended (Wise 18). Research has found that relatively few school districts have successful evaluation systems that communicate to teachers where improvement is needed; therefore, schools need to develop new, more comprehensive programs (Wise 20). McGreal states that systems built around attitudes directed toward improving teacher performance increase the teachers' positive feelings toward evaluation and increase the possibility that teachers will change their behavior (6). Groups of teaching behaviors have been shown by research to be related effective teaching (White 90).

Although school districts nationwide differ greatly in size, staff, student diversity, and goals, several factors have been identified by research to contribute to the success of a teacher improvement evaluation system. By identifying these common factors, school systems can use evaluation to improve teacher performance resulting in more positive teacher reaction.

STATEMENT OF THE PROBLEM

Discovering and using common factors found in successful evaluation systems can aid school districts in the development of evaluation programs designed to improve teacher performance or effectiveness.

LIMITATIONS OF THE STUDY

It is impossible to dictate one single method of evaluation for every school because each school is unique in its goals, values, and customs, depending on the community that surrounds it. Each system must develop its own evaluation program after examining its own goals, priorities, and curriculum. Although evaluation systems vary from school to school, many common factors have been found which make these programs successful. This study examines the various methods of evaluation and seeks to find those factors necessary for enhancing teacher effectiveness.

In addition, everything that occurs in a classroom cannot be measured by an evaluative instrument. Much of learning and teacher/student interaction depends on the rapport established between teachers and their students. Plans that school districts develop interpret the criteria and add descriptors to fit their own needs and goals.

DEFINITION OF TERMS

Career development plan - training and practice over a period of time to improve teaching skills (Hanes 11). Cited 25

Classroom management - competencies in student monitoring and organization of the lessons and assignments (Manatt, Palmer, Hidlebaugh 22). Cited 14

Classroom performance - competencies directly related to instructional delivery (Schlechty 16). Cited 25

Evaluation feedback - teachers are provided with information on their performance after evaluation (Duke 32). Cited 6

Faculty performance - competencies in work-related activity outside the classroom (Schlechty 16). Cited 26

Formative evaluation - evaluation system used to promote faculty instructional development for growth and increasing professionalism (McGreal 2). Cited 8

Observation - collection of information by actual viewing of a classroom and teacher's performance (Lawton 33). Cited 6

Professional performance - competencies that teachers have or do to enhance their abilities (Schlechty 16). Cited 26

Rater bias - the influence of factors such as life style and behavior attributes rather than performance on the evaluation rating (Bullock 9). Cited 9

Reliability of evaluation - the consistant correlation of evaluation ratings from one evaluation to the next (Savage 4). Cited 7

Self-evaluation - teachers analyze their own performances using a rating scale, videotape, or audiotape (Aubrecht 417). Cited 8

Student evaluation - measurement instruments where students rate teachers (Manatt 10). Cited 8

Student performance evaluation - pupil achievement on standardized tests used to evaluate teacher performance (Medley 179). Cited 8

Summative evaluation - evaluation system designed to be used for making personnel decisions and school accountability (McGreal 2). Cited 8

Teacher evaluation - a system used by a school district to define the teaching task and provide a mechanism for judging teacher performance (Wise 23). Cited 6

Teacher mentors - experienced teachers who work with new or probationary teachers to improve skills (Conley 12). Cited 26

Validity of evaluation - actual relationship between what is observed and the results of the investigation so that the results accurately reflect the performance (Conley 61). Cited 10

CHAPTER II

OUTLINE

I. Status of Teacher Evaluation
   A. Teachers' feelings
      1. No effect on teaching
      2. Incompetent evaluators
   B. Supervision versus evaluation
      1. Plan of growth
      2. Rating
   C. Types of evaluation
      1. Summative
      2. Formative
      3. Models
      4. Key ingredients

II. Problems with Typical Evaluation Methods
   A. Principal as observer
      1. Inconsistant
      2. Biased
   B. Student evaluations
      1. Forms used
      2. Use with other methods
   C. Self-evaluation
      1. Modify behavior when discrepencies occur
      2. Self-analysis methods
   D. Student achievement
      1. Teacher influence
      2. Teacher control over environment
      3. Regression effect
   E. Rating scales
      1. Non-teaching criteria
      2. Evaluator training
      3. Feedback
   F. Observation
      1. Feedback
      2. Time and number
      3. Criteria based on research
      4. Evaluator training
   G. Teacher involvement
      1. Selection and implementation of system
      2. Peer evaluation and coaching

III. Attributes of Successful Evaluation Programs
   A. Committment
   B. Purpose
   C. Teacher involvement
   D. Levels
   E. Data base

IV. Case Studies of Successful Research-Based Programs

ABSTRACTS (arranged as used in outline)

Turner, R.R. "What Teachers Think about Their Evaluations." Educational Digest 52 (Feb., 1987): 40-43.

One thousand teachers responded to a poll in May and June of 1986 for Learning 86. In this poll, teachers gave their views on an ideal evaluation. The teachers stated that evaluators (to be effective) should be interested in the evaluation process and be a common sight in the classroom. They spent ample time in observation and knew the classroom and students well. Of the respondents, 49% said their teaching improved. Twelve percent said that a significant improvement was noted. Fifty-two percent, however, said that evaluation had a negative effect or no effect on their teaching. These feelings were caused by evaluators who based their judgments on one fifteen minute evaluation per year or who were more concerned with neat rooms or fancy bulletin boards than teaching criteria. The greatest failing of evaluators was the failure to provide feedback, suggestions, or reasons for ratings to the teacher. Fifty-three percent of the teachers said that their last evaluation was an accurate assessment, in their opinion, of their teaching; however, areas where improvement could be made were not noted. Principals were the evaluators in 72% of the cases. Many teachers preferred a team evaluation rather than a single individual. They felt a team would be more effective in identifying incompetent teachers. Sixty-seven percent of the teachers knew of incompetent teachers in their school systems.

Duell, Orpha K., and Ron Davison. "Elementary Principals' and Teachers' Views of Various Modes of Instructional Evaluation." ERS Spectrum 5 (Winter, 1987): 23-33.

Duell and Davison conducted a study in which questionaires were sent to 100 elementary teachers and 100 elementary principals from midwestern school districts. The purpose of the study was to determine the viewpoints of school personnel concerning traditional modes of evaluation. Forty-nine items were on the questionnaire. The teachers and principals rated each item on a six to one scale. Discrepencies between the teachers' and principals' mean responses were identified by computing two-tailed t tests between the ratings for each item. Teachers rated observation by the principals with a follow-up conference as accurate (4.31 on the six point scale); however, it was tenth on their list beneath such items as self-evaluation, student responsiveness, student gains in knowledge, and informal discussions with teachers. On the other hand, principals rated observation with a conference second (4.69) in their list of accurate measurements. This finding was significant at the 0.01 level with a t of 2.68 showing that principals viewed observation with a follow-up conference as more accurate than did teachers. Both teachers and principals felt that self-evaluation, student responses, and observation were accurate means of evaluation. Observation had to be followed by a conference to be accurate. Teachers and principals agreed that the least accurate modes were those of which the teacher was unaware (listening on the intercom) or which gave them no feedback. Teachers and principals felt most comfortable with the methods they felt were most accurate.

McCarty, Donald J., Johanna W. Kaufman, and Julie C. Stanford. "Supervision and Evaluation - Two Irreconcilable Processes?" The Clearinghouse (April, 1986): 351-353.

The authors defined supervision as a developmental process of directing, guiding and supporting the instructional behavior of teachers to enhance the quality of instruction offered to students. The primary aim of supervision was not a formal rating of worth, but rather was guided by peer-defined standards of practice. Evaluation was defined as a performance judgment about the effectiveness and quality of teaching designed to rank teachers on a continuum. In a study of 76 teachers in 36 school systems in Wisconsin, the authors determined teachers' views on the evaluation process and its effects on their teaching. Eighty percent of the teachers reported that evaluation consisted of one in-class visit every two or three years followed by a global formal rating. This activity was seen as superficial and useless for determining quality teaching. Reasons given for the feeling of futility included the use of rating scale evaluation forms where all teachers were rated highly, and no real areas for improvement were noted. When supervision was introduced into the evaluation process through a clincal method or teaching model, its effectiveness depended on the teachers' perception of how competent the principals/evaluators were and how well they led the implementation of the program. Teachers seldom mentioned that suggestions from principals changed their teaching. Rather, motivation for change came from students, peers, or themselves.

Burke, Peter J., and Robert Kray. "Experienced and Inexperienced Teachers Attitudes Toward Supervision and Evaluation." Report. 66th Annual Meeting. Chicago, IL: American Educational Research Association, 1985.

Fifty experienced teachers and 50 inexperienced teachers were surveyed by the authors to determine their views on actual evaluation experiences and their attitudes toward preferred practices. The researchers devised six statements outlining supervisory tasks and six statements describing evaluative tasks to which respondents agreed or disagreed to on a five point Likert-type scale. The reliability of the scale was determined to be 0.90 for the supervisory items and 0.790 for the evaluation items. The original hypothesis of the study was that experienced teachers would differ significantly in their attitudes about evaluation and supervision from inexperienced teachers. Two sample t tests were used to determine any significant differences in the mean responses of the two teacher groups as well as each group's attitudes about real evaluation experiences and ideal ones. Comparing the real and ideal scales for both groups, no real significant differences were found except for one item for the experienced teachers. They wanted to know how they compared to other teachers, but in real situations they did not receive this information. This item, interestingly, was an evaluation item. Both groups tended to favor supervisory items over evaluation items overall. Experienced teachers could not agree on what type of practices were ideal. Both groups agreed that supervision encouraged and supported, not judged or restrained.

McGreal, Thomas L. Successful Teacher Evaluation. Alexandria, VA: Association for Supervision and Curriculum Development, 1983.

After examining evaluation systems in schools nationwide, McGreal presented the view that summative evaluation was used for making personnel decisions or weeding out the bad teachers while formative evaluation was similar to supervision in that the goal was faculty development and improvement. He contended that an evaluation system could not be built on either type alone, but should include both. Highly summative systems were found to involve little or no encouragement for teachers to improve instructional behavior. Instructional improvement models were shown to improve the likelihood for change in teacher behavior. Five types or models of evaluation were found. Sixty-five percent of United States school districts used the Common Law Model where the emphasis was on high supervisor-low teacher involvement in observation for summative purposes. The standardized criteria imposed the same standards on all teachers. Seventy percent of the criteria were administrative (cooperation, attendance) rather than teaching oriented. The Goal Setting Model, however, focused on teacher growth and improvement by having the teachers work toward goals set by themselves. Observation was used to measure progress toward those goals, but took extensive amounts of time due to the number of conferences necessary for feedback. The Product Model used student test results to measure teacher performance; however, norm-referenced standard tests and the regression effect often gave inaccurate measures. The Clincal Supervision Model involved observation and analyzation of the teacher's classroom behavior for improvement. The new Artistic Model promoted judging teaching as an art illicting expressive outcomes from students. Seven ways to evaluate that McGreal listed were observation, parent evaluation, peer evaluation, student performance, self-evaluation, student evaluation, and artifact (tests, handouts) collection.

Duke, Daniel, and Richard Stiggins. Teacher Evaluation - 5 Keys to Growth. AASA, NAESP, NASSP, NEA, 1986.

The authors' research consisted of examining case studies from 30 teachers who had experienced positive growth as a result of effective evaluation. Common dimensions of each evaluation were compared to identify active ingredients to success. Several hundred teachers were then asked on a questionnaire which dimensions were present in their last evaluation. Five key ingredients were identified from this study. First, and most critical, were the attributes of teachers. Openness to suggestions, orientation to change, instructional competence, and subject knowledge were among the significant attributes necessary. Second, evaluators or observers showed credibility, persuasiveness, patience, and trust. Evaluators relied on department chairpersons for additional information on the subject area. Ample time was allowed for observation so evaluators had familiarity with the teacher and classroom. Another key factor was the evaluation data. Criteria were well defined and clearly communicated to teachers. Developmental criteria were used in addition to summative criteria. Feedback was an essential key to promote teacher improvement. Suggestions for improvement, courses of action, and the performance related to the pre-set standards needed to be communicated. The setting or context where evaluation took place was the last key. Evaluation related to state law, district policy, and union contracts and policies. A good evaluation system allowed time and resources for teacher growth.

Medley, Donald M., and Homer Coker. "The Accuracy of Principals' Judgments of Teacher Performance." Journal of Educational Research 80 (March/April, 1987): 242-247.

This study was designed to test the accuracy of principals' judgments of the performance of the teachers they supervised. Also studied was the accuracy of the judgment of performance depending on which principal did the observation, the grade level the teachers taught, the ability of the students, and the achievement gains as measured by standardized tests in math and reading. Forty-six principals and 322 teachers participated. Correlation coefficients were calculated separately for 87 groups of teachers, and estimated student gains were adjusted to allow for the regression artifact. Each class was also split into random halves, and correlation coefficients were estimated twice. All the principals used a standard form to record their judgments. The results of the study showed that principals rated teachers in their own schools much higher than those of other schools. The mean correlation over all items on the form was 0.17. The accuracy of principals' judgments showed no significant difference depending on which principal made the judgment. Judgments for odd numbered grades appeared to be more accurate than those from even numbered grades. From these results, the authors concluded that the average principal is not a good judge of teacher performance based on the low accuracy statistics found. This study showed that objective and accurate observations were far more difficult to obtain than earlier research had shown.

Bullock, William Jr., and Joseph Davis. "Interpersonal Factors That Influence Principals' Rating of Teacher Performance." Planning and Changing 16 (Spring, 1985): 3-11.

Thirty-nine elementary and high school principals and 290 teachers participated in research to determine whether principals' rating of teachers was related to the principal-teacher life style and interpersonal needs compatibility. The Life Style Orientation Questionnaire was used to determine whether individuals fit into formalistic, sociocentric, or personalistic life styles. The fundamental Interpersonal Relations Orientation Behavior Scale determined need compatibility based on inclusion, control, and affection. For each dimension, subscales of wanted behaviors and expressed behaviors were calculated. The last form used in this study was a teacher rating checklist designed for this research. Correlations were found to be significant at the 0.01 level. Responses to two questionnaires on needs and life style were scored to check the compatibility of the pairs of principals and teachers. Scores were then correlated with the principals' ratings of teachers' performances. The research revealed that principals were significantly influenced by attraction factors and rated teachers with whom they shared life values and behavior attitudes higher. When compatible in individual freedom, group orientation, rules, laws, policies, and order, principals rated teachers higher. Principals gave higher ratings to teachers with whom their interpersonal needs were satisfied. The most stable relationship was shown to be between those who differed in respect to dominance and submission. The study concluded that the principals' evaluative responses were biased by life style and need compatibility.

Shrock, Sharon A., and Others. "Instructional Evaluation Techniques: Ends Dictate Means." Report. Annual Conference. Philadelphia, PA: Educational Communications and Technology, 1987.

Shrock's study involved the use of different student evaluation forms used to rate professors at the university level. Two samples of data were taken, one in the spring of 1980 and another in the fall of 1980. Frey's evaluation form was used; it was designed to receive student input on the professor's administrative decision making. It used a strongly agreed to strongly disagree scale to measure such items as planning, discussion, presentation of material, and accuracy of grading. Two summary scores of rapport and pedagogical skill were available on the Frey form. The second instrument used was the Southern Illinois University Instructor Course Evaluation Form (ICE) which contained 40 items clustered under the subtitles Personal-Interpersonal, Course Structure, Course Quality, and Course Difficulty. These two forms showed different types of correlations. On the Frey scale, the rapport and pedagogical skill did not influence each other (correlation 0.19); whereas, on the other form, all four subscales were related and dependent on each other. The first sample included 13 classes and 11 professors, and the second sample had 17 classes and 11 instructors. The results showed that due to the relationships in the ICE form, two different outcomes resulted depending on the form used. This study recommended that these types of instruments be used for diagnostic, not adminsistrative uses. The authors stated that student evaluation was useful, but it should be accompanied by observation by trained evaluators.

Savage, Tom V., and Marsha K. McCord. "The Use of Student Evaluation in the Assessment of Teacher Competence." Report. 67th Annual Meeting. San Francisco, CA: American Educational Research Association, 1986.

The authors first reviewed other research to determine whether student evaluation was a valid method of teacher evaluation. Although teachers and principals tended to mistrust student evaluation, the research from several sources showed that it was a valuable addition to the other methods. The stability of student ratings from one year to the next showed correlations of 0.87 to 0.89. Other methods were also used to show validity such as alumni ratings of teachers that were very similar to the rates of students currently enrolled. Research showed that student evaluations were best for the interactive aspects of teaching. The mean response was found to be the best for interpretation. The authors then studied the effect of student evaluation data on a group of 51 principals. Summative evaluations of seven teachers were given to each principal. Then one group of principals was given student evaluations classified as high. A second group was given student evaluations classified as low. Another group was given only the summative data. No significant difference among the three groups was found at the 0.05 level. The principals perceived the students' evaluations as not a valuable source of information even though studies had shown otherwise. Although student evaluations were proven to be valid by research, this study showed the principals did not see it as such and were not likely to use the data as an additional evaluative tool.

Aubrecht, Judith D., Gerald S. Hanna, and Donald P. Hoyt. "High School Student Ratings of Teaching Effectiveness Compared with Teacher Self-Ratings." Educational and Psychological Measurement 46 (Summer, 1986): 415-423.

Aubrecht and her associates studied the correlations of students' and teachers' responses on the same evaluative instrument to determine whether student evaluations of teachers were valid. Over 500 teachers were used in the sample. The sample involved 11 school districts chosen as a rough representation of the national population. Two classes for each teacher were used to provide data. The instrument used was the IDEA Form H. Students rated their own progress toward nine general instructional goals, while teachers completed the same form for both of their classes from the teacher's perspective. Results from four separate analyses were compiled. In examining student-teacher agreement by item, modest correlations of 0.20 for median values were found where values of 0.10 or above were considered significant. The median r for ratings of progress toward educational goals was 0.30. Overall student-teacher agreement on relative ratings of items showed r's of 0.90 to 0.69. Thus, the relative ratings were quite similiar; however, absolute teacher and student means showed differences by item type. For progress items, the mean difference was 0.04; for environmental items, the difference was 0.19; and for behavior items, the teachers rated themselves higher with a mean difference of 0.38. Within class student-teacher agreement on relative rating of items showed positive mean correlations with the median r value of 0.55. Student-teacher agreement in discriminating between class ratings had a median value of 0.26. The four analyses provided evidence for the validity of high school student ratings of teacher instruction.

Moses, Ingrid. "Self and Student Evaluation of Academic Staff." Assessment and Evaluation in Higher Education 11 (Spring, 1986): 76-86.

The study examined the results of self evaluation by teachers in the Australian university setting followed by comparison with student ratings. Instructors completed a questionnaire designed to serve as an aid to instructional decision making. Correlations were studied between the self evaluations and student evaluations after three evaluative rounds. The overall effectiveness rating was shown on a scale of one to seven with one being poor and seven outstanding. The students rated staff higher (5.23) than did the staff themselves (4.87). The correlation was 0.07 showing that the two variables were unrelated, nor were any relationships shown for evaluations over 6.0 or under 4.5. Instructors did not perceive themselves as the students did. Moses then examined two specific groups: those teachers who rated themselves significantly higher or lower than the student ratings. Those instructors who rated themselves highly came from departments where good teaching had been rewarded, and they were very confident. Those who rated themselves lower adjusted their self evaluations upward. The author concluded that self evaluation alone or student evaluation alone did not change teaching behavior. Change occurred when discrepancies were found between the two ratings. Thus, self evaluation should be accompanied by student evaluations, and comparisons made of the two ratings.

Freiberg, H. Jerome. "Teacher Self-Evaluation and Principal Supervision." NASSP Bulletin 71 (April, 1987): 85-92.

The author compared several means of assessing teacher effectiveness as well as providing a means of feedback for teacher change. Student gains on standardized tests provided information at the end of the year often too late to institute changes for the current group. Teacher-made tests provided information as well, but the teacher often did not know where changes in teaching strategy needed to be made. Observation feedback was the most prevelant means of providing data; however, time was short and principals did not offer specific suggestions in most cases. Peer feedback was found to be rare in most schools. Teacher self-evaluation was found to be helpful if used in conjunction with other means of supervision. The author developed the Low Inference Self-Assessment Measure (LISAM) which gave a teacher the opportunity to audiotape and then analyze lessons. Teachers identified whether their lessons included higher cognitive questions, whether teacher-student talk was balanced, whether a motivating set and closure were present, and whether enough wait time was present after each question. Positive statements and the use of student ideas were also determined. The author identified the conditions necessary for success with this type of self-assessment. The climate for success or failure was supported by the administration. Teachers discussed the results of the LISAM with the principals and made changes based on the data. Finally, teachers saw a clear benefit from the assessment and were supported by staff development activities.

Medley, Donald M., Homer Coker, and Robert S. Soar. "The Inadequacy of Existing Methods of Evaluation." Teacher Competence. Bloomington, IN: Phi Delta Kappan, 1985.

The reliability of using student standardized achievement test scores was questioned as a basis for teacher evaluations. Several points were presented against the use of such tests. Correlations of 0.40 to 0.70 were found between IQ and measures of aptitude and achievement. When evaluations were based on the class mean score, the correlations were as high was 0.90 indicating that as much as 80% of the achievement was already present before teachers influenced it. Standardized tests also held teachers accountable for pre-existing conditions over which they had no control. The home condition had a correlation of 0.47 with test scores in some cases. Peer group pressure, whether from high achievers or low achievers, also influenced test scores. Another point of contention was the stability of the tests when used to measure student gains. Research showed stability coefficients of 0.30; however, the authors stated that coefficients of 0.90 were needed to be usable to predict teacher effectiveness. Using the Spearman-Brown formula, a mean of 20 scores with a reliability of 0.30 would compile a reliability of 0.90. Only after twenty years could competency be determined by this method. The regression effect of achievement tests was the most important reason against their use. The top 10% of students and the bottom 10% of students tended to have scores which regressed toward the mean. Negative correlations of -0.30 to -0.50 were found for these groups. Thus, if students were assigned to teachers on the basis of test scores, the only teachers that were evaluated fairly were teachers assigned to the middle group of students.

Haertel, Edward. "The Valid Use of Student Performance Measures for Teacher Evaluation." Education, Evaluation, and Policy Analysis 8 (Spring, 1986): 45-60.

Haertel surveyed the research concerning the use of standardized achievement test scores and other student performance data as means of evaluating teacher performance. Several recommendations were made before student performance could be used for evaluation. Achievement tests were to be referenced to curricular or instructional objectives. The more generalized the test was, the less it would measure teacher competence. Students were pretested and posttested to determine gain in a specific area. This helped to eliminate the use of previously gained knowledge of students which would distort the findings. Adjustments were made for factors beyond the teacher's control such as home environment, absenteeism, and school climate. Student scores were accompanied by periodic observations, and teachers kept portfolios of examples of student work and progress. The author foresaw problems in using student scores including teachers who altered their instruction to teach test material and gave practice tests over tested materials. Problem solving and discovery learning might be eliminated from classroom work. To make evaluation comparisons of teachers, comparable aptitudes of students, learning objectives, and school resources were needed. Students needed incentives to perform well especially if teachers were disliked. Pilot studies were needed to determine norms for teacher comparisons. Teachers not meeting the norms were observed and student portfolios examined. Using student scores was both expensive and time consuming because special tests, pretesting, posttesting, and moderately complicated statistical analysis were needed.

Manatt, Richard P., Kenneth L. Palmer, and Everett Hidlebaugh. "Evaluating Teacher Performance with Improved Rating Scales." NASSP Bulletin. 60 (Sept, 1976): 21-25.

Five years of research by the education department of Iowa State University sought to develop improved rating scales for use in evaluation. The research yielded a 30 item rating scale used in a five point Likert-type responses. These thirty items were analyzed for reliability and discriminating power by using the Menne and Tolsma F test. Sixty-nine elementary and secondary teachers were evaluated by students, peers, administrators, and the subjects themselves to determine these 30 items from a total of 360 performance descriptions. The items were placed on the scale in order of discriminating power under the subtopics of productive teaching techniques, positive interpersonal relations, organized/structured class management, intellectual stimulation, and desirable out-of-class behavior. Preliminary norming of the scale showed discrimination between teachers of high, medium, and low performance. The research recommended a pilot year to establish district norms based on local appraisor skill, state laws, and district policies. The recommended use of the scale included self-appraisal using the scale, a pre-observation conference to discuss objectives and methods, two or three classroom observations, post observation conferences to discuss findings and note progress, agreement on a plan of action to improve, time and help to improve, and a final summary evaluation.

Tobin, Kenneth. "Validating Teacher Performance Measures Against Student Engagement and Achievement in Middle School Science." Science Education 70 (Oct., 1986): 539-547.

The Teacher Performance Instruments (TPAI), developed by the Georgia State Department of Education, had proven to be a valid measure of classroom management and student engagement and achievement in science classes in three other studies. Tobin selected 13 teachers from four middle schools in Clarke County, Georgia to further study the TPAI rating scale. Students' reasoning ability was tested using the Test of Logical Thinking which had an internal consistency reliability of 0.85. Students were randomly selected using the test score to obtain a broad range of formal reasoning ability. Pupil achievement was measured by a test consisting of two items for each of twelve objectives with an alpha coefficient of 0.8. The classroom procedures and interpersonal sections of the TPAI were used to evaluate teachers on competencies such as planning and management, indicators which were criteria to measure competencies, and descriptors to determine if specified behaviors occurred. The rating scale was between one and five for four observations of an eight lesson sequence. The results of the research were correlations of 0.77 between total teacher performance and average class rates of engagement and 0.48 with total score average achievement. Reliability of the total TPAI measure was 0.87. Average summative achievement was 35% with a standard deviation of 8%. Tobin concluded that higher teacher performance on the TPAI was associated with higher rates of student engagement and higher process skill achievement.

Johnson, Nancy C., and J. Kenneth Orso. "Teacher Evaluation Criteria." ERS Spectrum 4 (Summer, 1986): 33-36.

Forty-eight large school systems from 48 states participated in a study of evaluation instruments. The authors analyzed the forms from these schools to compile a group of criteria and categories which were common to all. In the responding schools, criteria were grouped into five basic catagories. The two most important and most prevelant were criteria concerning instruction and classroom management. Three other categories focusing on non-instructional criteria included professional responsibilities, personal characteristics, and interpersonal relations. The categories were found by the authors to be what teachers and administrators perceived to be qualities that made good teachers. The authors also discovered that other factors made evaluation successful. Evaluators needed to be properly trained in the use of the instrument. Time for proper evaluation was essential, and the purpose of the evaluation needed to be made clear to all. Teachers who were involved in the development of the evaluation instruments and procedures were better informed about the purposes and were more committed to the process. Communication about the use of evaluation for decisions concerning tenure, retention, and teacher improvement was also crucial.

McIntyre, Kenneth E. "Using Classroom Observation Data for Diagnostic Purposes." Report. 67th Annual Meeting. San Francisco, CA: American Educational Research Association, 1986.

McIntyre compared the use of observation as an evaluative measure with efficiency scores based on an input-output model developed by the Educational Productivity Council of the University of Texas at Austin. When selecting an observation instrument, the focus was on teacher and student behavior in a descriptive manner. The evaluators were thoroughly trained and observed by taking into account differences in teacher performance from situation to situation. The instrument revealed teacher objectives, learning activities, classroom management, and rewards for desirable behavior. Observers recognized that "more" was not necessarily better when observing teacher behavior. The instrument developed by the University of Texas included only those indicators or criteria supported by process-product research. In this study, three evaluators made two full-period visits to classrooms of 24 teachers. One hundred forty visits were made using 31 tested indicators. The input-output criteria on a one to seven scale focused the observers' responses on specific behaviors. An analysis was done to the scores of each observer to arrive at a total score for each teacher. Using this method, weaknesses in specific areas were found and provided data to be used to improve teacher performance. The author concluded that an analysis of efficiency scores from an input-output model assisted observers in comparing teachers' performances and identifying areas where teachers needed to improve.

Selmes, Cyril. "Teacher Evaluation in the Classroom." Educational Management and Administration 14 (Autumn, 1986): 191-196.

Research in the United Kingdom and United States revealed that some assumptions held about evaluation were not always effective such as "correct" ways of teaching were recognized by "good" teachers. Another incorrect assumption was that evaluation should be hierarchical because evaluators recognized correct teaching methods. Research revealed that the vast majority of teachers were competent, and that evaluation should be more concerned with teacher improvement rather than dismissal of an incompetent few. Research stated that the hierarchical system of evaluation was what teachers objected to most. To avoid these objections, the prime purpose of evaluation was staff development carried out by the teacher's peers, fellow teachers or professionals. The evaluation was based on criteria jointly negotiated and agreed to by teachers. The evaluation system also provided data about teaching that could be discussed and from which improvements were planned. Studies had shown several ways to meet these suggestions. The use of videotape with discussion before and after compared the theory and actual practice of the teacher. Teachers needed to identify qualities to be evaluated. These criteria were placed on an evaluation sheet to be used by the teacher and the evaluator. The videotape and sheets were used in the last stage of evaluation which discussed the findings and made a plan for changes to improve the quality of teaching. Two advantages were listed for this method. One was not imposing criteria on professionals, but giving those professionals voices in the process, and the second was the evaluation sheet which identified differences recognized by peers and professionals. As with any evaluation system, this required time to explain and negotiate.

Lawton, S.S., and Others. Development and Use of Performance Appraisal of Certified Education Staff in Ontario School Boards. Toronto, Canada: Ontario Department of Education, 1986.

Thirty school districts in Ontario were surveyed to determine what types of performance appraisals were used, the extent of implementation, the most effective practices, and the processes for development and implementation. One thousand principals, superintendents, and directors responded, and 4092 teachers responded. Two types of appraisal were found: appraisal designed to bring about change in performance (formative) and appraisal designed to judge performance (summative). Before appraisal, teachers were usually notified in person, and a pre-conference was held in many districts to set objectives and goals. The main obstacle to not having a pre-conference in some districts was lack of time for scheduling one. In the survey, 70% of the teachers felt the purpose of evaluation was to decide permanent contracts and then to improve instruction. The criteria used in the evaluation seemed to fit the purposes because they were related to classroom techniques, management, and teacher-student relations. The criteria were incorporated most often into a rating scale used during observation. Ninety-six percent of the evaluations involved observation with teacher input and information giving. In 89% of the cases, observation was done by one person. The emphasis in the province seemed to be, however, on observation, not follow-up. Lack of time prevented post-conferences. Only 21% reported follow-up conferences which averaged 21 minutes in length. Only 14% of the teachers reported the development of a plan of improvement from evaluation. Thus, many teachers did not perceive improvement as a result of the appraisal. After statistical analysis of the study's responses, the following characteristics were found to have positive correlations to favorable outcomes of evaluation:

1. Purposes were stated clearly in advance.
2. Classroom performance was the major criterion.
3. Follow-up conference was held with the development of a plan for improvement.
4. Teachers were involved in the development and implementation of the policies.

White, Kinnard, Marvin D. Wyne, Gary B. Stuck, and Richard H. Coop. "Assessing Teacher Performance Using an Observation Instrument Based on Research Findings." NASSP Bulletin 71 (March, 1987): 89-95.

The authors examined over 600 research studies to determine which teaching practices were consistently chosen by research as being related to student achievement. Twenty-five practices found most often were arranged to develop an observation instrument to be used in teacher evaluation. Five broad teaching functions were identified including management of instructional time, management of student behavior, instructional presentation, instructional monitoring, and instructional feedback. The instrument was then evaluated by a panel of nationally recognized authorities in effective teaching research and tested by principals in a variety of schools. Two uses for the instrument were found. First, it was used during observation to provide data on which to make plans for improving teaching skills. It helped focus the observer's attention on specific teaching practices known to be related to student achievement. Second, the instrument was used for personnel appraisal. It was recommended that only the five teaching functions be used for this purpose rather than all twenty-five criteria. Reliability was tested for this method by using videotape. When principals were trained in the use of the instrument, reliabilities of 0.82 were found. The authors felt that by using research methods to develop evaluation means, a scientific basis was established for guiding professional practice.

Edwards, Clarence M. Jr. "An 'Effective Teaching' Approach to Teacher Evaluation and Staff Development." ERS Spectrum 4 (Spring, 1986): 3-8.

Edwards headed a study group in Orange County, Virginia that concluded that evaluation should concentrate on improving the effectiveness of the 98% of competent teachers rather than trying to identify and remove the incompetent 2%. The study group began by identifying more than 100 performance indicators grouped under twelve general categories. After two years of observation and documentation, the indicators were narrowed to 15 which seemed to be present in each successful instructional presentation. Each teacher was then given the opportunity to participate in a professional growth program over a period of five years. During this time, teachers were observed 30 times to demonstrate proficiency in all 15 performance areas. Classroom teachers were allowed to choose the indicator to be observed and the class the observer would document. Observers followed three stages including a pre-conference, observation for strengths to satisfy performance indicator, and post conference. Although the principals were the chief overseers and observers, classroom teachers were allowed release time to observe and coach other teachers. If observations did not show competence, teachers were guided in methods to adjust teaching strategies. Teachers were surveyed annually to determine their reaction to the plan. Eighty-four percent felt the system accurately reflected their teaching performance, and 81% felt their teaching had improved.

Shapiro, Phyllis, Phyllis Cohen, Nancy Weight, and Gloria Pollock. "Supervision - A new Approach to Inservice Teacher Education." Unpublished paper, 1979.

A suburban elementary school outside of Toronto was used by the authors to study the supervision method of evaluation using a Tuckman Form modified to their purposes. The Tuckman Form measured classroom climate which was the responsibility of the teacher to develop. Classroom climate could either promote or inhibit learning. The Form consisted of 28 pairs of adjectives which were bipolar such as cold/warm, patient/impatient, disorganized/organized. Four clusters of classroom climate and teacher behavior were determined by factor analysis. These were creativity, dynamism, organized demeanor, and warmth or acceptance. Reliability ranged from 0.55 to 0.84 for internal analysis and 0.65 to 0.90 for inter-rate coefficients. The implementation of the form followed a six step formative approach. First, a good relationship between supervisors and the teachers was promoted. Second, a plan for observation was developed followed by observation where both teachers and observers filled out the form. Then scoring was completed by the research staff and a graph was prepared for the fifth stage which was feedback to the teachers. Any disagreement in rating areas was seen as a consideration for teacher improvement or change. The last stage involved planning follow-up activities and the next cycle of evaluation to monitor any changes. The authors concluded that this method provided positive feedback for teachers as well as a means of involving them in the process.

Ashbaugh, Carl R., and Katherine L. Kasten. "Should Teachers Be Involved in Teacher Appraisal?" NASSP Bulletin. 71 (Sept., 1987): 50-53.

In a review of early and current research on the subject of teacher involvement in evaluation, the authors stated that studies supported the concept of teacher involvement. Administrators tended to disagree on the amount of involvement teachers should have. Two tests were recommended to determine if teacher participation was acceptable. First, the issue or decision was determined to have had a high personal stake, be important, or relevant. Second, if the teachers had knowledge or experience in the decision area, they were involved. Evaluation was determined to meet both of these tests in the formative area. Many teachers were reported to feel uncomfortable when judging peers summatively. Studies recommended that the time to involve a committee of teachers was in helping to select an evaluation procedure. Administrations then needed to develop expertise in all teachers of the system with training for the selected system. Finally, teachers were involved in their own formative evaluation and the monitoring of their peers rather than summative evaluations of them.

Christen, William L., and Thomas J. Murphy. "Inservice Training and Peer Evaluation: An Integrated Program for Faculty Development." NASSP Bulletin 71 (Sept., 1987): 10-18.

In developing one of the career ladder plans in Arizona, interviews were conducted with educators and private businessmen to determine which concepts were important to staff growth and development. One concept that was recommended was career-long inservice training about effective teaching practices and strategies followed by practice and experimentation with new techniques. The second concept was a support group of peer-evaluators to provide feedback and dialogue between teachers. This allowed the teachers to experiment with different methods without fear of dismissal in summative evaluation. Teachers felt peer evaluation was more acceptable and accurate than that of administrators. Peer evaluation was used to provide a formative evaluation with the goal of improving teacher performance. Administrators then evaluated summatively for decisions in personnel matters. Feedback sessions were essential to the peer evaluation process where what actually occurred was discussed and changes were planned for observation in the next session. The authors concluded that this two step process treated teachers as individuals and brought about lasting changes in teacher performance.

Harris, Douglas E., and Keith Pillsbury. "Shared Governance of Staff Evaluation." Educational Leadership 44 (April, 1987): 62.

The Essex Town, Vermont, School District developed a review committee to oversee their chosen evaluation system which used both formative and summative evaluation. The review committee was designed to monitor the system for flexibility, appraisal of function, changes needed, and shared responsibility of teachers and administrators. The membership of the committee consisted of two teachers per building, a counselor, a special educator, a subject-area coordinator, a principal, and an assistant superintendent. The committee gathered information from staff members to assess the system, recommended changes, explored new evaluative techniques, introduced new staff members to the system, shared the system with other groups, and conducted formative and summative reviews of the system. The authors found that when teachers and administrators shared the responsibility for successful evaluation, everyone was more accountable, and the system was more responsive to the needs of the staff.

Buttram, Joan L., and Bruce L. Wilson. "Promising Trends in Teacher Evaluation." Educational Leadership 44 (April, 1987): 5-6.

The authors reviewed current literature on teacher evaluation for Research for Better Schools. Five trends had become evident in recent years. First, the school districts were developing more comprehensive evaluation systems based on teacher effectiveness literature. Studies showed that the research based criteria correlated with student engagement rates, achievement gains and perceptions of the school climate or learning environment. Next, research showed that evaluators were being trained to standardize procedures using supervised practice with videotape and live classroom observations. The third trend was a move toward accountability by making evaluations important and appropriating time and money for them. Also, many school districts were using the evaluation results as a means to train teachers to be more effective such as prescribing formal course work, holding inservice sessions, or pairing successful and unsuccessful teachers for individual assistance. Finally, the administration and the teachers union in many districts collaborated to supervise first year teachers, develop new evaluation plans, and provide follow-up staff development. Teachers became partners in the selection, use, and follow-up of the evaluation system.

Conley, David T. "Critical Attributes of Effective Evaluation Systems." Educational Leadership 44 (April, 1987): 60-64.

Conley, a senior consultant for the Colorado Department of Evaluation, analyzed the evaluation systems of Colorado's 177 school districts. Eight aspects surfaced as being essential to a successful, effective evaluation system. As the system was introduced, all participants accepted the validity of the system based on district values and goals. All participants were then trained in the mechanics of the system. The teachers were aware of the evaluation procedures and expectations. Evaluatees needed to know the rationale and relative weight of the performance criteria. They then set their goals based on district priorities. Also evaluators were trained in the use of the system including observation, analysis, conferencing, goal setting, and remediation for ineffective teachers. An important aspect was the use of different levels of evaluation other than just summative levels for personnel decisions. Effective systems added an improvement and growth track for improvement of skills and a remedial track for those teachers with serious problems and in danger of dismissal. These levels of evaluation were separate entities. The formative evaluation and growth evaluation consisted of the largest portion of the assessment cycle. The summative portion for decision making was used later to observe improvement made. Successful systems employed a variety of evaluation methods such as teacher goal setting, clinical supervision, observation, and student achievement which were used to develop individual potential. Finally, the district considered evaluation a top priority by allotting time, resources, and training for it.

Wise, A.E., L. Darling-Hammond, M.W. McLaughlin, and H.T. Berstein. Teacher Evaluation: A Study of Effective Practices. Santa Monica, CA: Rand Corp., 1984.

This study began with a preliminary survey of 32 school districts designated as having highly developed evaluation systems. An intensive research of four case study districts followed. Teacher satisfaction with evaluation was found to be related to the perception that all evaluators shared the same criteria, frequent samples of performance, frequent communication and feedback, and the ability to affect and have some control over the assessment. While systems differed in instrumentation, frequency of evaluation, and teacher roles, several common factors were found to make evaluation successful. Each case study district showed a commitment to evaluation, had procedures for ensuring evaluator competence, collaborated with the teacher organization and individual teacher, and had compatibility with district goals and strategies. Each case study had four basic purposes for evaluation including teacher improvement, accountability, organizational decision making, and personnel decisions. The most important purpose was felt to be teacher improvement which was developed through a process including pre-conferences, classroom observation, then follow-up conferences. This study developed five conclusions which have guided many other research projects. They were as follows:

1. Evaluation matched the goals of education of the district and community. The school system examined its goals and adopted an evaluation system compatible with them.
2. The level of commitment to and resources for evaluation outweighed checklists and procedures. Most importantly, time was allowed; regular assessment and evaluator training were also important.
3. The school district decided the main purpose of evaluation and then matched the process to the purpose.
4. To sustain resource commitments and political support, teacher evaluation was seen to have utility and achieved reliability, validity, and cost effectiveness.
5. Teacher involvement and responsibility improved the quality of teacher evaluation. Teachers helped select, implement, and oversee the process as well as advise and coach their peers.

Leach, J. William, and Lester Solomon. "Performance-Based Certification in Georgia: Present and Future." Report. Washington, D.C.: National Commission on Excellence in Teacher Education, 1984.

The Georgia Teacher Certification Tests were designed to be job-related and to reflect the minimum content knowledge necessary to teach in a certified classroom area. Beginning teachers were assessed first by a Teacher Certification Test which determined whether the person had the knowledge to teach in his respective teaching field. The Teacher Performance Assessment Instrument (TPAI) was then used in on-the-job assessment of actual teaching. Field testing had proven this instrument to be reliable and valid. It was composed of 14 competencies each with two to five performance indicators rated on a one to five scale during observation. To receive a permanent teaching certificate, beginning teachers had to demonstrate proficiency in all 14 competencies within three years or six assessment opportunities. An advancement in salary also accompanied the demonstration of proficiency. A regional three member team then reviewed the TPAI scores and teachers' portfolios and interviewed teachers as well as observed them to determine the final assessment score for advancement. The majority of teachers took more than one assessment to demonstrate mastery. Seventy-five percent successfully completed the process in two assessments; the remaining teachers were given additional opportunities. The assessment proved successful with 90% of the participants having positive responses about the assessment and its feedback to them.

Capie, William, S.J. Anderson, C.E. Johnson, and C.D. Ellett. Teacher Performance Assessment Instrument: A Handbook for Interpretation. Atlanta, GA: Georgia State Department of Education, 1979.

The handbook for the Georiga TPAI instruments was designed to aid the evaluator in assessing teacher performance. The instrument contained five parts, three of which were used for certification. The Teacher Plans and Materials Instrument focused on a portfolio for a unit prepared by the teacher. These materials were discussed in the teacher interview and observation collected data about their usefulness. The Classroom Procedures Instrument focused on classroom practices as determined by observation. The social atmosphere of the classroom and warmth and friendliness were measured by the Interpersonal Skills Instrument also based on observation. The Professional Standards Instrument and the Student Perceptions Instrument were used for professional development and research. Three assessors were used with each observation including a data collector, a master teacher, and a department head or principal. Each person did an independent rating which was combined with the others and graphed onto a performance profile to determine proficiency. This author was also involved in several studies establishing the validity (content, discriminate, construct, and criterion-related) and reliability of the TPAI instruments.

Smith, B. Othanel, Donovan Peterson, and Theodore Micceri. "Evaluation and Professional Improvement Aspects of the Florida Performance Measurement System." Educational Leadership 44 (April, 1987): 16-19.

Developers of the Florida Performance Measurement System spent six years in testing the summative and formative instruments used to assess teacher performance in Florida. Six domains of teaching behavior were evaluated. They included planning, management of student conduct, organization and development of instruction, presentation of subject matter, communication, both verbal and non-verbal, and testing of students and feedback to them. The summative instrument was used to screen for problem areas; the formative instrument was used to pinpoint specific behaviors needing remediation. The instruments were thoroughly researched and tested before applied. The reliability coefficients ranged from 0.79 to 0.98. In 1,223 observations for norming, no significant differences were found based on sex, race, experience, class enrollment, or socio-economic status. There were significant differences between post-elementary and elementary grades so two norming groups were developed. The validity studies showed consistent positive relationships between instrument scores and student engagement and achievement. Factor analysis supported the domain structure and the equal weighting and scoring of items. Observers were trained in an intensive three-day program followed by testing in the competence areas and research. The observer recorded only the occurrence of the teaching behavior, not its value. Thus the Florida System indicated to teachers where they were strong or weak in performance areas.

Peaker, Geoffrey. "Teacher Management and Appraisal in Two School Systems in the Southern U.S.A." Journal of Education for Teaching 12 (1986): 77-83.

Peaker, a member of the British Schools Inspectorate, visited, analyzed, and compared the evaluation systems of two school systems, Dekalb County, Georgia and Palm Beach County, Florida. These findings were related to systems studied in Britain. He found that the major purpose of the assessment in each system was improvement in education for accountability of the school system. Evaluation was also used to determine promotions or incompetent teachers in need of remediation. Each system used three appraisers with instrumentation based on research data identifying specific teacher behaviors which correlated with student learning. Data collectors were extensively trained in the three areas of appraisal: portfolio on preparation, interview of teacher, and observation using a research based instrument. Peaker indicated that the instruments emphasized behaviors which were measured by a change in pupil performance. More difficult and complex behaviors and teacher student relationships were not assessed because they could not be measured. This emphasis on measurement led to teaching methods that were highly controlled or "safe" rather than flexible and innovative. The procedures did not take into account varying school conditions or teachers' wider contributions to the school and classroom. What could be measured was highly valued; what could not be measured was overlooked.

Tesch, Stephanie, Larry Nyland, and Dough Kernutt. "Teacher Evaluation - Shared Power Working." Educational Leadership 44 (April, 1987): 26-30.

Using McGreal's (1983) goal-based appraisal process, the Pasco, Washington School District developed an evaluation system over a four year period. The system was designed to help teachers develop appropriate teacher improvement goals, make observation a more comfortable, formative process, and establish mutually agreed upon minimum expectations for teachers. One teacher and three administrators formed the planning committee which researched and developed the pilot program. The pilot plan was tested in two ways. Half of the building principals used it on all their teachers; the other half of the principals tested on two of their best teachers. After the pilot tests proved successful, the teachers were divided into three groups. Track A (satisfactory teachers) were extensively observed and set goals for instructional improvement which were observed and monitored in the evaluation process. Tract B (satisfactory teachers) were monitored for minimal requirements and switched places with the Track A teachers the following year. Track C included beginning or remedial teachers who were given extensive appraisal and guidance to improve their teaching skills. Those teachers demonstrating success moved to Track A or B; those who did not improve moved to the legal process of probation or non-renewal. In a January, 1986 survey, both teachers and principals reacted positively to the goal-setting and guidance process.

Furtwengler, Carol B. "Tennessee's Career Ladder Plan: They Said It Couldn't Be Done." Educational Leadership 43 (Nov., 1985): 50-56.

The Tennessee Interim Certification Commission spent one year reviewing literature and feedback from teachers to develop the criteria for the Tennessee Career Ladder Program. The program was based on improvement of teacher performance as well as increased responsibility within the system with the result of student gain and monetary reward. Six major areas of competence were identified including preparation for instruction, appropriate use of teaching strategies for content and objectives, use of evaluation for improvement, management of classroom activities, establishment of a professional leadership role, and effective communication. Multiple data sources were used to evaluate each teacher for these criteria. Sources included classroom observation, a teacher-prepared portfolio, interview, peer questionnaire, principal questionnaire, student questionnaire, and a written test. Three evaluators participated in the pre- and post-conferences and observation using a coding system to record amounts of actual behavior. The evaluators were higher level teachers specifically trained in the use of the procedures. After evaluation, teachers moved up a five step ladder from probationary to apprentice teacher, through Career Level I, II, and III. Career III teachers had longer contracts, higher pay, and served as peer evaluators and department heads, and taught additional classes in summer sessions.

Furtwengler, Carol B. "Lessons from Tennessee's Career Ladder Program." Educational Leadership 44 (April, 1987): 67-69.

After three years of use and study of the Tennessee Career Ladder Program, several changes were made to make the system more usable and efficient. Although teachers were involved in the initial planning stages of the program, communication at the local levels was felt to be weak. Programs of meetings, videotapes, workshops, and hotlines were instituted. A Teacher Orientation Manual was published to lessen the communication gap. Another issue was the quantitative scoring which indicated frequency of behavior rather than quality. Scoring was changed to be more qualitive and was made available and explained to teachers prior to evaluation. Portfolio preparation and extensive interviews were found to require excessive teacher time which did not aid in classroom instruction. The portfolio was reduced from a total instructional plan to a report of professional and leadership activities. The interviews became one hour dialogues where each evaluator discussed a major criterion area. The Tennessee Instructional Model was developed to use in training sessions for the evaluative process. This also gave Career Level II and III teachers a means of assisting beginning teachers. Each teacher who was evaluated by the program also assessed it using a one to five scale. Teachers rated classroom observation highest among the seven data sources giving it a 3.8 rating. The multidata source system received a rating of 4.2 and peer evaluators a 4.3 rating. The original program was very time consuming and stressful for many teachers, but after modifying the program, positive results were seen, and teachers were eager to participate.

Hanes, Robert C., and Kay F. Mitchell. "Teacher Career Development in Charlotte-Mecklenburg." Educational Leadership 43 (Nov., 1985): 11-13.

The authors examined specific aspects of the Charlotte-Mecklenburg Plan to determine their effectiveness. Vital to the program were the two balanced parts, the action-growth plan and the observation of performance by peer observers. Teachers were helped to achieve competencies by the use of feedback from the two areas. The teacher mentor played a key role in this process. Rapport, trust, confronting feedback, and quality time were factors involved in a working relationship between mentor and teacher. Beginning teachers were surveyed to determine the most valuable aspects of working with teacher mentors. Areas rated highest included informal conversation, 67%; management of student behavior, 27%; instructional presentation and content, 19%; and management of instructional time, 14 1/2%. Another area reviewed was the administrative role involving principals and assistant principals. Principals felt that their role was strengthened by the multiple sources of data and the direction the conferences and plans gave. The last area examined was the action-growth plan which concentrated on the skills necessary for effective classroom performance. The teacher and evaluative team set goals, identified strategies to achieve those goals, and listed observable results the observation team saw and measured. The career development plan helped many teachers to expand their teaching materials and techniques to improve student learning.

Schlechty, Phillip C. "Evaluation Procedures in the Charlotte-Mecklenburg Career Ladder Plan." Educational Leadership 43 (Nov., 1985): 14-19

Literature and studies of personnel evaluation in high technology and information processing industries served as a basis for the Charlotte-Mecklenburg Career Ladder Plan rather than educational research. Through summative and formative evaluation, the system was designed to communicate performance expectations, provide feedback, provide information on needed teacher training, provide data to assess the quality of teacher training, and make personnel decisions. The industrial research method proved to be very similar to educational research in its methods and principles. Evaluation was continuous, was based on multiple criteria, used multiple evaluators, had data collected over a period of time for personnel decisions, and was evaluated by teachers. Classroom performance, faculty performance, and professional performance were assessed using a team of three evaluators including a senior teacher mentor. Teachers progressed through Career Level I, II, or III by demonstrating competency in each area. A series of announced and unannounced observations were made in teachers' classrooms with conferences emphasizing goal setting to meet the competencies following. After the competencies were met, the teacher prepared an action-growth plan outlining performance goals. The teacher's plan included the action, results, assessment method, and criteria to be observed. Thus, both the observation report and the action-growth plan were part of the evaluation and provided a source and direction for the teachers' improvement.

Manatt, Richard P. "Lessons from a Comprehensive Performance Appraisal Project." Educational Leadership 44 (April, 1987): 8-14.

The Research Institute for Studies in Education was asked by five school districts in adjoining areas in Minnesota and Iowa to research and develop teacher evaluation plans. Five years were spent in developing the models used by the group of districts. Teachers formed the basis for the planning phase where the emphasis was placed on learner outcomes, effective teachers in effective schools, training for performance, and feedback for all participants. In the School Improvement Model (SIM), several sources of data were related statistically by computer including curriculum mapping plans, criterion and norm-referenced student testing twice a year, evaluations of all participants (teachers, principals, and boards), student ratings of teachers, school climate surveys, job improvement targets, and attendance records. Different sets of criteria were used for teachers, principals, and boards which were subjected to reliability and validity tests before implementation. All criteria were discriminating at the 0.01 or 0.05 levels. Reliability on the many instruments ranged from 0.597 to 0.91. A Behaviorally Anchored Rating Scale was used to determine the extent to which employees exhibited the behavior or performed the criterion increasing scatter so all scores were not superior ratings. The researchers found that formative observation and evaluation took a great deal of time; whereas, summative evaluation took six to eight hours to complete. Formative recommendations were two announced visits with pre- and post-conferences and several drop-in visits. The computer-based comparison allowed reports of staff development needs and performance to be analyzed over the district for goal and growth decisions.

Guidelines for Performance-Based Teacher Evaluation in Missouri. Jefferson City, MO: Missouri Department of Elementary and Secondary Education, 1984.

The evaluation plan researched and developed in Missouri contained four general performance areas assessing the instructional process, classroom management, interpersonal relationships, and professional responsibilities. Nineteen specific criteria based on research were selected for the four areas with descriptors under each to provide specific guidelines during evaluation. The system was designed to reflect measurable, observable teacher behavior in order to improve classroom instruction and increase student learning. Building principals were responsible for the assessment process and were trained in the methods of analyzing teacher behavior and the use of the process. The evaluation process began with a formative phase involving pre-conferences and scheduled observations using the designated criteria. Following the observations, post-conferences were held for feedback to the teachers. After discussing strengths and weaknesses, job-target forms were filled out outlining measurable goals and objectives that teachers were to work toward. Additional personnel assisted teachers in attaining the targets if necessary. The targets remained in effect until satisfactorily completed as observed in scheduled and unscheduled visits by the evaluators. At the end of the evaluative cycle, the evaluators completed summative reports which reviewed all pertinent data about teachers' performances. A final conference was held to review the summative report. If teachers were not performing at an acceptable level, then the improvement plans were continued.

CHAPTER III

SUMMARY

Many evaluation systems used in schools were inadequate and failed to provide sufficient feedback to the teachers for improvement of performance. As many as 52% of teachers interviewed felt that evaluation had a negative effect or no effect on their teaching (Turner 40). These comments resulted from poorly trained evaluators, lack of defined criteria, and lack of suggestions or directives for improvement. Teachers viewed observation as less accurate and helpful than principals did (Duell 24). Both groups felt that observation with a follow-up conference was more accurate than observation alone. The least accurate methods were those that the teacher was unaware of or that gave no feedback (Duell 28).

Evaluation which involved supervision as well as observation was rated higher by teachers depending on their perception of the competence of the evaluator (McCarty 352). Supervision was seen as a plan of growth and improvement guided by the evaluator or peers. Motivation for change in teaching most often came from students, peers, or teachers themselves (McCarty 353). Teachers often agreed that supervision encouraged improvement but could not define ideal supervisory practices (Burke 8).

Two types of evaluation were present in research. Formative evaluation (supervision) was designed to help teachers improve their teaching performance through feedback and guidance; whereas, summative evaluation was a final assessment of teachers' performances used for personnel decisions. Successful systems of evaluation included both (McGreal 2). Five types of models were described in research: The Common Law Model, the Goal Setting Model, the Product Model, the Clinical Supervision Model, and the Artistic Model (McGreal

9). Methods of evaluation included observation, parent evaluation, peer evaluation, student performance, self-evaluation, student evaluation, and artifact collection (McGreal 125-128).

Five key ingredients were shown by research to help teachers experience positive growth as a result of evaluation. Teachers were open to suggestions and willing to change. Evaluators were credible and allowed ample time for evaluation. Evaluation criteria were well defined and communicated to the teacher. Essential to teacher improvement were feedback and development of strategies for better performance. The context of evaluation was based on district policy and teacher involvement (Duke 14).

Principals were found by research to be inaccurate in their assessments of teacher behavior. They tended to rate their own building's teachers higher than they did other teachers (Medley and Coker 245). Principals were also biased by the relationship of life style and need compatibility to the teachers they evaluated (Bullock 10).

Student evaluations used to assess teacher performance did not give an accurate assessment of teachers in some research studies. Different evaluative scores resulted from using different forms (Shrock 4). Other studies showed that student evaluations were valid for measuring interactive aspects of teaching (Savage 4). Research recommended that student evaluations be used as supplemental material for formative evaluation rather than the only source of assessment scores (Shrock 13). Teacher self-evaluations, in addition to student evaluations, were helpful to encourage teachers to modify behavior when discrepancies were shown (Moses 83). Analysis of lessons using audiotape to measure frequency of behavior allowed teachers to make a self-analysis of their lessons and make changes for improvement (Freiberg 91).

Research showed several problems concerning the use of student performance on standardized tests as a means of evaluation. As much as 80% of the achievement shown on the tests was already present before current teachers influenced it. Home environment and peer pressure, factors that teachers could not control, were shown to influence scores (Medley, Coker, and Soar 179). The regression effect distorted achievement test scores so that only the middle group of scores reflected accurate measurements of achievement (Medley, Coker, and Soar 181). If student performance scores were used for evaluation of teacher effectiveness, special tests which matched teachers' objectives were more valid than general achievement tests. Pre-testing and post-testing were recommended to accurately determine student gain. Complicated statistical analysis was also necessary to adjust for factors beyond the teachers' control (Haertel 59).

Rating scales often were based on non-teaching criteria (McGreal 14). Research had shown that rating scales were useful for evaluation if based on research data and used in conjunction with observation and pre- and post-conferences before a final, summative evaluation was made (Manatt, Palmer, and Hidebaugh 23-24). Rating scales based on research showed high correlations between high teacher performance and higher student engagement and achievement (Tobin 445). For rating scales to be effective, the evaluator needed to be trained in the use of the instrument, sufficient time needed to be allowed to do a thorough analysis, and the results needed to be communicated to the teacher (Johnson 36).

Observation was one of the most common means of evaluation using either peers or an administrator. Evaluators needed guidelines when observing to make the observation valid whether those guidelines were input-output models

(McIntire 7) or research based teacher effectiveness criteria (Johnson 33). For observation to be useful for teacher improvement, several conditions were necessary. Teachers were more willing to discuss and improve after observation if they had been allowed to help develop the criteria and instrument (Selmes 192). Post conferences with feedback and the development of an improvement plan were necessary for the teacher to improve after observation (Lawton 36). The amount of time spent in observation and follow-up was found by research to be a crucial factor in improving effective teacher behavior (Lawton 35). The criteria used as the basis for observation was most effective when it was based on research. Most instruments examined used the basic teaching functions of classroom management, instructional time management, instructional presentation, instructional monitoring, and instructional feedback (White 91). Training of the evaluator in the use of the instrument made the observation more reliable (White 94).

Peer evaluation was seen as an effective means of monitoring behavior and performance especially when peers had expertise and knowledge of the subject area evaluated (Ashbaugh 51). Peers were also helpful in the formative process to provide information and feedback about effective performance (Christen 13).

Although many types of evaluation methods and instruments were researched, no one system was found to be the best. Several attributes were found to be present in successful assessment systems. First, the school district showed a definite commitment to evaluation, and the instrument used was compatible with the district's goals. Teacher improvement was seen as the major purpose developed through formative procedures. Time, resources, and training were committed to the process to make it useable, valid, and reliable. Teachers were involved in the evaluation process at all levels from development and

implementation to peer evaluation and coaching (Wise 13-15). Successful systems also used different levels of evaluation (formative and summative) and a multisource data system to provide input from a variety of levels within the district (Conley 63-64).

Successful evaluation systems based on research and designed to meet the goals of the districts that developed them include the Georgia TPAI, the Florida Performance Measurement System, the Pasco, Washington School District system, the Tennessee Career Ladder Program, the Charlotte-Mecklenburg Career Ladder Plan, the School Improvement Model, and the Missouri Performance-Based Teacher Evaluation program. These systems were developed after extensive research and testing which proved the chosen criteria valid and reliable. Their main purpose was teacher improvement after a long period of formative evaluation. Teachers were involved in the development and implementation of these systems and received extensive feedback about their performance from them.

CONCLUSIONS

Evaluation can help teachers experience positive growth and improvement. Research has shown that no one method of evaluation is best. The evaluatee, evaluator, feedback, criteria, and context are all key ingredients to a successful assessment system. Most methods of evaluation have inaccuracies and disadvantages when used alone or without sufficient communication with the teacher. Successful systems have been shown by research to have several attributes in common. The evaluation plan matches the goals of the school district, and the district commits sufficient time and resources to the chosen plan. The chosen criteria in successful plans are based on research of

effective teaching methods, and the use of a multi-source data program gives a total view of the teacher's performance. Teachers feel more positive about evaluation if the evaluator is thoroughly trained in the use of the instruments, and gives feedback which the teachers can use to alter their classroom behavior. Research shows that formative evaluation helps teachers change performance by setting goals or developing a plan for more effective teaching. After a cycle of formative evaluation, summative evaluation is used to give a final rating of effectivenss or for personnel decisions. Instruments and criteria are tested before implementation to insure validity and reliability. Teachers feel the evaluation program is successful and effective if they participate in all phases of its selection, implementation, and use. Evaluation systems using these precepts are viewed positively by teachers and administrators and are valid and reliable.

RECOMMENDATIONS

School systems which are reassessing their evaluation systems or developing new systems for accountability, teacher improvement, and personnel decisions should follow these recommendations:

1. First, examine goals and priorities before selecting an evaluation system. Match the chosen plan with the directives of the district.

2. Allow time and resources to make evaluation useful at all levels. Time for several evaluations and feedback conferences is especially valuable.

3. Base criteria for evaluation on effective teaching research and measure the domains of classroom management, instructional presentation, and instructional time management. The district may add other domains as well.

4. Use more than one source of information gathering to gain a broader base for evaluation. Peers, students, and the teachers themselves can provide valuable input.

5. Train evaluators in the use of the chosen instrument(s). Feedback is essential from evaluators to teachers so that teachers can change classroom techniques.

6. Include both formative and summative evaluation in the program. Teachers need to know strengths and weaknesses and be given time to improve before a final rating is given.

7. Test all parts of the plan before the program is implemented to insure validity and reliability.

8. Involve teachers in the selection, implementation, and use of the program. Involved teachers will be more willing to change their behavior and become more effective in the classroom.

WORKS CITED

Ashbaugh, Carl R., and Katherine L. Kasten. "Should Teachers Be Involved in Teacher Appraisal?" NASSP Bulletin 71 (Sept., 1987): 50-53. (Cited 18, 31)

Aubrecht, Judith D., Gerald S. Hanna, and Donald P. Hoyt. "High School Student Ratings of Teaching Effectiveness Compared with Teacher Self-Ratings." Educational and Psychological Measurement 46 (Summer, 1986): 415-423. (Cited 11)

Bullock, William Jr., and Joseph Davis. "Interpersonal Factors That Influence Principals' Rating of Teacher Performance." Planning and Changing 16 (Spring, 1985): 3-11. (Cited 9, 29)

Burke, Peter J., and Robert Kray. "Experienced and Inexperienced Teachers Attitudes Toward Supervision and Evaluation." Report. 66th Annual Meeting. Chicago, IL: American Educational Research Association, 1985. (Cited 7, 28)

Buttram, Joan L., and Bruce L. Wilson. "Promising Trends in Teacher Evaluation." Educational Leadership 44 (April, 1987): 5-6. (Cited 20)

Capie, William, S.J. Anderson, C.E. Johnson, and C.D. Ellett. Teacher Performance Assessment Instruments: A Handbook for Interpretation Atlanta, GA: Georgia State Department of Education, 1979. (Cited 22)

Christen, William L., and Thomas J. Murphy. "Inservice Training and Peer Evaluation: An Integrated Program for Faculty Development." NASSP Bulletin 71 (Sept., 1987): 10-18. (Cited 19, 31)

Conley, David T. "Critical Attibutes of Effective Evaluation Systems." Educational Leadership 44 (April, 1987): 60-64. (Cited 20, 32)

Duell, Orpha K., and Ron Davison. "Elementary Principals' and Teachers' Views of Various Modes of Instructional Evaluation." ERA Spectrum 5 (Winter, 1987): 23-33. (Cited 6, 28)

Duke, Daniel, and Richard Stiggins. Teacher Evaluation - 5 Keys to Growth AASA, NAESP, NASSP, NEA, 1986. (Cited 8, 29)

Edwards, Clarence M., Jr. "An 'Effective Teaching' Approach to Teacher Evaluation and Staff Development." ERS Spectrum 4 (Spring, 1986): 3-8. (Cited 17)

Frieberg, H. Jerome. "Teacher Self-Evaluation and Principal Supervision." NASSP Bulletin 71 (April, 1987): 85-92. (Cited 12, 29)

Furtwengler, Carol B. "Lessons from Tennessee's Career Ladder Program." Educational Leadership 44 (April, 1987): 67-69. (Cited 25)

Furtwengler, Carol B. "Tennessee's Career Ladder Plan: They Said It Couldn't Be Done." Educational Leadership 43 (Nov., 1985): 50-56. (Cited 24)

Guidelines for Performance-Based Teacher Evaluation in Missouri Jefferson City, MO: Missouri Department of Elementary and Secondary Education, 1984. (Cited 27)

Haertel, Edward. "The Valid Use of Student Performance Measures for Teacher Evaluation." Education, Evaluation and Policy Analysis 8 (Spring, 1986): 45-60. (Cited 13, 30)

Hanes, Robert C., and Kay F. Mitchell. "Teacher Career Development in Charlotte-Mecklenberg." Educational Leadership 43 (Nov., 1985): 11-13. (Cited 25)

Harris, Douglas E., and Keith A. Pillsbury. "Shared Governance of Staff Evaluation." Educational Leadership 44 (April, 1987): 62. (Cited 19)

Johnson, Nancy C., and J. Kenneth Orso. "Teacher Evaluation Criteria." ERS Spectrum 4 (Summer, 1986): 33-36. (Cited 14, 30, 31)

Lawton, S.B., and Others. Development and Use of Performance Appraisal of Certified Education Staff in Ontario School Boards Toronto, Canada: Ontario Department of Education, 1986. (Cited 16, 31)

Leach, J. William, and Lester Solomon. "Performance-Based Certification in Georgia: Present and Future." Report. Washington, D.C.: National Commission on Excellence in Teacher Education, 1984. (Cited 22)

Manatt, Richard P. "Lesson from a Comprehensive Performance Appraisal Project." Educational Leadership 44 (April, 1987): 8-14. (Cited 26)

Manatt, Richard P., Kenneth Palmer, and Everett Hidlebaugh. "Evaluating Teacher Performance with Improving Rating Scales." NASSP Bulletin 60 (Sept., 1976): 21-25. (Cited 13, 30)

McCarty, Donald J., Johanna W. Kaufman, and Julie C. Stanford. "Supervision and Evaluation - Two Irreconcilable Processes?" The Clearing House (April, 1986): 351-353. (Cited 1, 7, 28)

McGreal, Thomas L. Successful Teacher Evaluation Alexandria, VA: Association for Supervision and Curriculum Development, 1983. (Cited 1, 8, 28, 29, 30)

McIntyre, Kenneth E. "Using Classroom Observation Data for Diagnostic Purposes." Report. 67th Annual Meeting. San Francisco, CA: American Educational Research Association, 1986. (Cited 15, 31)

Medley, Donald M., and Homer Coker. "The Accuracy of Principals' Judgments of Teacher Performance." Journal of Educational Research 80 (March/April, 1987): 242-247. (Cited 9, 29)

Medley, Donald M., Homer Coker, and Robert S. Soar. "The Inadequacy of Existing Methods of Evaluation." Teacher Competence Bloomington, IN: Phi Delta Kappan, 1985. (Cited 12, 30)

Moses, Ingrid. "Self and Student Evaluation of Academic Staff." Assessment and Evaluation in Higher Education 11 (Spring, 1986): 76-86. (Cited 11, 29)

Peaker, Geoffrey. "Teacher Management and Appraisal in Two School Systems in the Southern U.S.A." Journal of Education for Teaching 12 (1986): 77-83. (Cited 23)

Savage, Tom V., and Marsha K. McCord. "The Use of Student Evaluation in the Assessment of Teacher Competence." Report. 67th Annual Meeting, San Francisco, CA: American Educational Research Association, 1986. (Cited 10, 29)

Schlechty, Phillip C. "Evaluation Procedures in the Charlotte-Mecklenburg Career Ladder Plan." Educational Leadership 43 (Nov., 1985): 14-19. (Cited 26)

Selmes, Cyril. "Teacher Evaluation in the Classroom." Educational Management and Administration 14 (Autumn, 1986): 191-196. (Cited 15, 31)

Shapiro, Phyllis, Phyllis Cohen, Nancy Weight, and Gloria Pollock. "Supervision-A New Approach to Inservice Teacher Education." Unpublished paper, 1979. (Cited 18)

Shrock, Sharon A., and Others. "Instructional Evaluation Techniques: Ends Dictate the Means." Report. Annual Conference. Philadelphia, PA: Educational Communications and Technology, 1987. (Cited 10, 29)

Smith, B. Othanel, Donovan Peterson, and Theodore Micceri. "Evaluation and Professional Improvement Aspects of the Florida Performance Measurement System." Educational Leadership 44 (April, 1987): 16-19. (Cited 23)

Tesch, Stephanie, Larry Nyland, and Doug Kernutt. "Teacher Evaluation - Shared Power Working." Educational Leadership 44 (April, 1987): 26-30. (Cited 24)

Tobin, Kenneth. "Validating Teacher Performance Measures Against Student Engagement and Achievement in Middle School Science." Science Education 70 (Oct., 1986): 539-547. (Cited 14, 30)

Turner, R.R. "What Teachers Think About Their Evaluations." Educational Digest 52 (Feb., 1987): 40-43. (Cited 1, 6, 28)

White, Kinnard, Marvin D. Wyne, Gary B. Stuck, and Richard H. Coop. "Assessing Teacher Performance Using an Observation Instrument Based on Research Findings." NASSP Bulletin 71 (March, 1987): 89-95. (Cited 2, 17, 31)

Wise, A.E., L. Darling-Hammond, M.W. McLaughlin, and H.T. Bernstein. Teacher Evaluation: A Study of Effective Practices Santa Monica, CA: Rand Corp., 1984. (Cited 1, 21, 32)

STUDENT QUALIFICATIONS

Linda Redinger SS#310-52-7640
5859 E. 20th Road
Argos, IN 46501

I. Educational Background
   A. Culver High School, Culver, IN, 1965
   B. Purdue University, West Lafayette, IN
      1. B.S. in Vocational Home Economics Education, J1969
   C. I.U.S.B., South Bend, IN, 1985
      1. 34 hours completed toward M.S. and English minor as of Dec., 1987.
      2. Complete M.S. in May, 1988.
      3. Complete English minor by 1990

II. Employment Experience
   A. Argos Community Schools, Argos, IN
      1. Substitute teacher 1969-1985, 1987-1988
      2. Vocational Home Economics teacher - part time 1969-1970, 1974-1975
      3. Title I Elementary Aide, 1985-1986
      4. English teacher, grades 11 and 12, 1987-1988 (one semester, for teacher on maternity leave)
   B. Triton High School, Bourbon, IN
      1. Vocational Home Economics teacher, grades 7-12, 1986-1987 (one semester, for teacher on maternity leave)
      2. Substitute teacher, 2nd semester, 1987

III. Reasons and qualifications for undertaking this study

I have worked in two different school systems with two very different evaluation programs, one a simple rating system and the other a more comprehensive goal-setting program.

As the result of participation in group projects in other classes, I became very interested in different methods of evaluation. Evaluation was dreaded by many of the teachers with whom I worked. Many were afraid that their job was at stake if they did not perform well. Others felt that the evaluator was not competent to evaluate their field of teaching, and he gave no information on how they could improve.

This study was undertaken to try to discover ways that evaluation could help a teacher improve classroom performance rather than just rate that performance as good or poor. If teacher effectiveness increased, then student interest and achievement should increase also. This study examined relationships between teacher performance, student learning, and evaluation.

ACKNOWLEDGEMENTS

Wanneta Reichert (English teacher) - assistance with form of proposal and final report.

Linda Nifong (friend, Gibson Insurance Agency) - typist.

Nancy Miller (English teacher) - proofreading final draft.